

# An efficient, variational approximation of the best fitting multi-Bernoulli filter

Jason L. Williams National Security and ISR Division, Defence Science and Technology Organisation, Australia  
and School of Electrical and Electronic Engineering, University of Adelaide, Australia  
email Jason.Williams@dsto.defence.gov.au

**Abstract**—Recent derivations have shown that the full Bayes random finite set (RFS) filter is comprised of a linear combination of multi-Bernoulli distributions. The full filter is intractable as the number of terms in the linear combination grows exponentially with the number of targets. A highly desirable approximation would be to find the multi-Bernoulli distribution that is closest to the full distribution in some sense, such as the set Kullback-Leibler (KL) divergence. This paper proposes an efficient, approximate method for achieving this, minimising a commonly-employed upper bound to the set KL divergence. The resulting algorithm is a principled, tractable, RFS-based extension of the popular set joint probabilistic data association filter. Subsequently, a special case of the method is utilised to provide an efficient approximation of the minimum mean optimal sub-pattern assignment estimator. The performance of the proposed methods is demonstrated in challenging scenarios in which up to twenty targets come into close proximity.

## I. INTRODUCTION

Recently, parallel derivations have found conjugate prior forms for target tracking using unlabelled random finite sets (RFSs) [1], [2] and labelled RFSs [3]. In each case, the form of the exact filter is a linear combination of multi-Bernoulli (MB) distributions. The complexity of exact methods is problematic as the number of terms in the linear combination grows exponentially in the number of targets; this is the problem of data association.

The complexity is addressed in [2] by seeking a MB distribution that approximates the posterior. Observing that the linear combination may be viewed as a marginalisation over a latent association variable  $a \in \mathcal{A}$ , the methods proposed approximated the probability distribution of data association,  $p(a)$ . This resulted in two approaches, the first of which was closely related to joint probability data association (JPDA) [4] and joint integrated PDA (JIPDA) [4], [5], and the second of which was related to the MeMBe filter [6], [7]. These approaches each have disadvantages: the former is near-optimal when targets are well-spaced but fails when targets are closely-spaced (referred to as coalescence, and studied in detail in [2], [8]), while the latter is more robust, but exhibits lower performance when targets are well-spaced.

A compelling alternative would be to find the MB distribution which minimises the set Kullback-Leibler (KL) divergence from the exact distribution. The first contribution of this paper is an efficient, approximate method of finding the MB distribution that minimises the set KL divergence. The resulting algorithm is related to set JPDA (SJPDA) [9], but is a principled, RFS-based alternative that accommodates

uncertainty in the number of targets. Furthermore, the computational complexity of the proposed method is drastically lower than SJPDA, requiring solution of a network flow linear program (LP) where the number of variables is  $N|\mathcal{H}|$ , rather than an optimisation involving  $N!|\mathcal{A}|$  variables (where  $|\mathcal{A}| \approx (|\mathcal{H}|/N)^N$ ). Since the term *variational inference* is widely used to refer to statistical approximations based on optimisation (e.g., [10]), we refer to our method as the variational MB (VMB) filter.

The optimal sub-pattern assignment (OSPA) was introduced in [11] as a measure of the difference between two sets of points. If  $X = \{x_1, \dots, x_n\}$ ,  $Y = \{y_1, \dots, y_m\}$ , and  $n \geq m$ , OSPA is defined as

$$d_{\text{ospa}}(X, Y) \triangleq \left[ \frac{1}{n} \min_{\pi \in \Pi_n} \sum_{j=1}^m d_c(x_{\pi(j)}, y_j)^p + c^p(n-m) \right]^{\frac{1}{p}} \quad (1)$$

where  $d_c(x, y) \triangleq \min\{c, d(x, y)\}$ ,  $d(x, y)$  is a distance function on the single target state space, and  $c > 0$  is a real number indicating the cost of a target not having a corresponding estimate (or vice versa). If  $n < m$  then the OSPA distance is  $d_{\text{ospa}}(Y, X)$  following the definition above (thus reversing  $n$  and  $m$ ).  $\Pi_N$  is the set of permutation functions on  $I_N \triangleq \{1, \dots, N\}$ , i.e.,

$$\Pi_N = \{\pi : I_N \rightarrow I_N | i \neq j \Rightarrow \pi(i) \neq \pi(j)\}$$

Minimum mean OSPA (MMOSPA) estimation has been studied previously in works such as [12], and it was the motivation for the approximation used in SJPDA. However, previous works have not satisfactorily addressed the growth in complexity with the number of targets. The second contribution of this paper is an approximation of the MMOSPA estimator based on the VMB algorithm; we refer to the result as the variational MMOSPA (VMMOSPA) estimator.

## II. BACKGROUND

### A. Random finite sets and multi-Bernoulli filters

RFS distributions provide an integrated mathematical framework for addressing estimation problems in which states to be estimated and/or observations form sets [6]. The difference between random sets and random vectors is that realisations of a random set will generally contain different numbers of elements, and the ordering of the elements is meaningless. RFS distributions encode both uncertainty in cardinality and uncertainty in values into a single representation. Uncertainty of this form could be represented via a cardinality distribution  $p(n)$

and a series of conditional state distributions  $f(x_1, \dots, x_n|n)$ ; the relationship between the RFS distribution  $f(X)$  and these components is:

$$f(\{x_1, \dots, x_n\}) = p(n) \sum_{\pi \in \Pi_n} f(x_{\pi(1)}, \dots, x_{\pi(n)}|n) \quad (2)$$

where the sum over permutations ensures permutation invariance.

Working within the RFS framework has several advantages. One of these is the ability to rigorously define measures of distortion caused to an entire multi-object distribution  $f(X)$  when approximating it by another multi-object distribution  $g(X)$  via the set KL divergence:

$$D(f||g) = \int f(X) \log \frac{f(X)}{g(X)} \delta X \quad (3)$$

where the set integral is

$$\int v(X) \delta X \triangleq v(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \dots \int v(\{x_1, \dots, x_n\}) dx_1 \dots dx_n \quad (4)$$

This measure encompasses both changes to the distribution of cardinality, and changes to the target state distribution.

Another advantage of working within the RFS domain is the ability to formulate dynamics and observation models in terms of sets (which is natural for many tracking problems) and calculate predicted and posterior distributions of these. Two recent works [2], [3] proposed slightly different conjugate prior forms for common assumptions. This work is based in the form studied in [2], which makes the following modelling assumptions.

**Assumption 1.** *The multiple target state evolves according to the following time dynamics process:*

- Targets arrive at each time according to a non-homogeneous Poisson point process (PPP) with birth intensity  $\lambda_b(x)$ , independent of existing targets
- Targets depart according to iid Markovian processes; the survival probability in state  $x$  is  $P_s(x)$
- Target motion follows iid Markovian processes; the single-target transition PDF is  $f_{t|t-1}(x|x')$

**Assumption 2.** *The multiple target measurement process is as follows:*

- Each target may give rise to at most one measurement; probability of detection in state  $x$  is  $P_d(x)$
- Each measurement is the result of at most one target
- False alarm measurements arrive according to a non-homogeneous PPP with intensity  $\lambda_{fa}(z)$ , independent of targets and target-related measurements
- Each target-derived measurement is independent of all other targets and measurements conditioned on its corresponding target; the single target measurement likelihood is  $f(z|x)$

Under these assumptions, [2] proves that the following form is a conjugate prior, i.e., it is preserved by prediction and

update:

$$f(X) = \sum_{Y \subseteq X} f_{\text{ppp}}(Y) f_{\text{mbm}}(X - Y) \quad (5)$$

where  $f_{\text{ppp}}(X)$  is a PPP representing unknown targets,<sup>1</sup> with intensity  $\lambda_u(x)$ :

$$f_{\text{ppp}}(X) = \exp \left\{ - \int \lambda_u(x) dx \right\} \prod_{x \in X} \lambda_u(x) \quad (6)$$

and  $f_{\text{mbm}}(X)$  is a MB mixture, i.e., a linear combination of multi-Bernoulli distributions, of the form

$$f_{\text{mbm}}(X) = \sum_{a=(h_1, \dots, h_N) \in \mathcal{A}} w_a \sum_{\biguplus_{i=1}^N X_i = X} \prod_{i=1}^N f_{h_i}(X_i) \quad (7)$$

where the notation  $\biguplus_{i=1}^N X_i = X$  denotes that the sum is over all disjoint subsets  $X_1, \dots, X_N$  whose union is  $X$ . As derived in [2], the terms in the sum  $a \in \mathcal{A}$  correspond to different choices of data association (i.e., different choices of which groups of past measurements correspond to the same targets, referred to as *global association hypotheses*); we assume that  $w_a \geq 0$  and  $\sum_{a \in \mathcal{A}} w_a = 1$ . The Bernoulli distributions for each component are drawn from a common set of single-target hypotheses  $\mathcal{H}$  (although the hypotheses utilised by each component  $i \in \{1, \dots, N\}$  may be disjoint), where  $f_h(X_i)$  is given by:

$$f_h(X_i) = \begin{cases} 1 - r_h, & X_i = \emptyset \\ r_h f_h(x_i), & X_i = \{x_i\} \\ 0, & |X_i| > 1 \end{cases} \quad (8)$$

In our implementations, we further assume that  $f_h(x_i) = \mathcal{N}\{x_i; \mu_h, \Sigma_h\}$ , although the basic derivation can handle other representations such as particle filters. As shorthand, when we refer to a global association hypothesis  $a$ , we assume that  $a = (h_1, \dots, h_N)$ , and hence we also refer to the  $i$ -th constituent single-target event as  $h_i$ .

Expressions and algorithms for predicting and updating the PPP component and the MB mixture are provided in [2]. The greatest difficulty in implementing the filter is the number of components in the MB mixture. In this work, we focus on the problem of simplifying the MB mixture into a single MB distribution. Therefore, for the purpose of this paper, we disregard the PPP component and refer to the MB mixture as  $f(X)$ .

## B. Set joint probabilistic data association

The Bernoulli distribution (8) is practically similar to the model used in integrated probabilistic data association (IPDA) [13]. Likewise, the TOM-MeMBer/P algorithm proposed in [2] is practically similar to JIPDA [5]. The method which we propose in this work is related to KL set JPDA (KLSJPDA) and set JPDA (SJPDA) [9]. KLSJPDA seeks to minimise the KL divergence between a modified distribution  $f(\mathbf{X})$  (where

<sup>1</sup>The model does not assume that  $P_d = 1$  for targets just born. Accordingly, a proportion of the targets hypothesised to have arrived by the birth model will go undetected. These targets, which have *never* been detected, are referred to as *unknown* targets. See [2] for a proof of the result and further discussion.

$\mathbf{X} = (x_1, \dots, x_n)$  is the vector of the joint state of all targets, and the number of targets  $n$  is known) and a Gaussian approximation:

$$\int f(\mathbf{X}) \log \frac{f(\mathbf{X})}{\mathcal{N}\{\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\}} d\mathbf{X} \quad (9)$$

The objective (9) is minimised with respect to the modified distribution  $f(\mathbf{X})$  (permuting elements in such a manner that the KL divergence is reduced but the symmetrised distribution is unchanged), and  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  (fitting a Gaussian to the modified distribution). SJPDA replaces this objective with the trace of the covariance  $\boldsymbol{\Sigma}$ , which is related to the mean OSPA metric. The modification of  $f(\mathbf{X})$  is also limited such that a fixed permutation is used under each global association hypothesis.

In comparison to SJPDA/KLSJPDA, the method proposed in the section III has the following advantages:

- The motivation and derivation is based on minimising the set KL divergence between a fixed distribution  $f(X)$  and a varying approximation  $g(X)$ , rigorously incorporating uncertainty in both number of targets and target states
- The formulation naturally handles cases in which the number of targets present is uncertain, whereas KL-SJPDA/SJPDA assume that the number of targets is known
- The method naturally yields a relaxed optimisation involving a single variable for each hypothesis and each track (i.e., Bernoulli component) that can be solved efficiently in low order polynomial time, whereas SJPDA involves a distribution of permutations for every global association hypothesis, resulting in time that is exponential in the number of targets

The final difference is especially significant, as it essentially means that SJPDA (and even more so KLSJPDA) can only be applied to very small problems, whereas the method proposed scales well with problem size.

### III. VARIATIONAL MULTI-BERNOULLI FILTER

The problem of interest is defined below. We refer to the resulting algorithm as the best fitting MB (BFMB) filter, and the subsequent approximation (described in section III-A) as the variational MB (VMB) filter.

**Problem 1.** Find the MB distribution  $g(X)$  that minimises the KL divergence

$$\begin{aligned} \operatorname{argmin}_{[g_j]} \int f(X) \log \frac{f(X)}{g(X)} \delta X = \\ \operatorname{argmax}_{[g_j]} \int f(X) \log g(X) \delta X \end{aligned} \quad (10)$$

where  $g(X)$  is MB:

$$g(X) = \sum_{\mathfrak{J}_{j=1}^N} \prod_{X_j=X} g_j(X_j) \quad (11)$$

and the components  $g_j(X_j)$  are similar in form to (8).

The RHS of (10) is obtained by separating the log of the quotient into the difference of the logs, and observing that the first term is constant WRT the variables of minimisation.

Lemma 1 shows that the multi-target set integral for a MB distribution can be decomposed into a series of Bernoulli set integrals.

**Lemma 1.** Suppose  $f(X)$  is as defined in (7), and  $v(X)$  is an arbitrary set-valued function. Then

$$\begin{aligned} \int f(X) v(X) \delta X = \\ \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) v\left(\bigcup_{i=1}^N X_i\right) \delta X_1 \cdots \delta X_N \end{aligned} \quad (12)$$

where, as stated previously, we assume throughout that  $a = (h_1, \dots, h_N)$ .

*Proof:* We first prove a case where  $f$  is a convolution of two distributions  $f_1$  and  $f_2$ , i.e., that

$$\begin{aligned} \int \left( \sum_{Y \subseteq X} f_1(Y) f_2(X - Y) \right) v(X) \delta X \\ = \iint f_1(X) f_2(Y) v(X \cup Y) \delta X \delta Y \end{aligned} \quad (13)$$

Starting from the LHS expression and using the shorthand  $X_n = \{x_1, \dots, x_n\}$ , and  $I_n = \{1, \dots, n\}$ :

$$\begin{aligned} \text{LHS} &= \sum_{n=0}^{\infty} \frac{1}{n!} \int \sum_{Y \subseteq X_n} f_1(Y) f_2(X_n - Y) v(X_n) dx_1 \cdots dx_n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{I \subseteq I_n} \int f_1(\bigcup_{i \in I} \{x_i\}) \cdot \\ &\quad \cdot f_2(\bigcup_{i \in I_n - I} \{x_i\}) v(X_n) dx_1 \cdots dx_n \quad (14) \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^n \frac{1}{m!(n-m)!} \int f_1(\{x_1, \dots, x_m\}) \cdot \\ &\quad \cdot f_2(\{x_{m+1}, \dots, x_n\}) v(X_n) dx_1 \cdots dx_n \quad (15) \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{m!k!} \iint f_1(\{x_1, \dots, x_m\}) \cdot \\ &\quad \cdot f_2(\{x_{m+1}, \dots, x_{m+k}\}) v(\{x_1, \dots, x_{m+k}\}) \cdot \\ &\quad \cdot dx_1 \cdots dx_{m+k} \quad (16) \\ &= \text{RHS} \end{aligned}$$

where (14) replaces the sum of subsets of  $X_n = \{x_1, \dots, x_n\}$  with a sum of the index subsets of the elements, and observes that the variable of summation is no longer a variable of integration, allowing the sum and integral to be exchanged; (15) observes that the integral (14) is the same for all  $I$  with  $|I| = m$ , and that there are  $\frac{n!}{m!(n-m)!}$  ways of choosing  $I \subseteq I_n$  with  $|I| = m$ ; and (16) makes a change of variables, defining  $k = n - m$ .

Subsequently, the desired result can be obtained by applying the two distribution case  $(N - 1)$  times for each  $a \in \mathcal{A}$ . ■

**Corollary 1.** Let  $[X_i] \triangleq (X_1, \dots, X_N)$ . Suppose that an alternative definition of a set-valued function  $\tilde{v}$  satisfies  $\tilde{v}([X_i]) = v(X)$  for any  $[X_i]$  such that  $\bigcup_{i=1}^N X_i = X$ . Then (12) can be equivalently evaluated as:

$$\int f(X)v(X)\delta X = \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) \tilde{v}([X_i]) \delta X_1 \cdots \delta X_N \quad (17)$$

**Lemma 2.** Suppose  $\tilde{v}(X) \triangleq c_{|X|} v(X)$ . Then

$$\int f(X) \log \tilde{v}(X) \delta X = \sum_n f(n) \log c_n + \int f(X) \log v(X) \delta X \quad (18)$$

where  $f(n)$  is the cardinality distribution corresponding to  $f(X)$ .

The proof of lemma 2 simply separates the log of the product into the sum of logs and simplifies. With these preliminary results, we are now ready to state the first theorem, that the RFS KL divergence for the MB distributions (10) can be optimised by summing over assignments of Bernoulli components in the original distribution to Bernoulli components in the simplified distribution.

**Definition 1.** Denote by  $\Pi_N$  the set of complete permutation functions on  $I_N \triangleq \{1, \dots, N\}$ :

$$\Pi_N = \{\pi : I_N \rightarrow I_N | i \neq j \Rightarrow \pi(i) \neq \pi(j)\}$$

**Theorem 1.** The solution of the optimisation

$$\operatorname{argmax}_{[g_j]} \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) \cdot \log \tilde{g}([X_i]) \delta X_1 \cdots \delta X_N \quad (19)$$

where

$$\tilde{g}([X_i]) = \sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i) \quad (20)$$

is the same as the solution of problem 1.

*Proof:* The difference between (11) and (20) is that the former considers assignment of the elements of  $X$  to  $g_j(X_j)$ , whereas the latter considers assignment of the  $N \geq |X|$  zero- or one-element subsets of  $X$  to  $g_j(X_j)$ . Thus, if  $n = |X|$ , there will be  $(N-n)$  empty subsets  $X_i$  in any decomposition  $\bigsqcup_{i=1}^N X_i = X$ . These could be assigned to the remaining Bernoulli distributions (i.e., those that were not assigned to non-empty  $X_i$ ) in  $(N-n)!$  ways, so  $\tilde{g}([X_i]) = c_{|X|} g(X)$  where  $c_n = (N-n)!$ . By corollary 1 and lemma 2, the two objectives (10) and (19) are different by a constant (WRT  $[g_j]$ ), so that the solution(s) attaining the maxima are the same. ■

#### A. Approximate solution of BFMB

We propose an approximate solution of (19) based on minimisation of an upper bound of the true objective. The derivation effectively treats the correspondence between the underlying Bernoulli distribution  $f_{h_i}(X)$  and the Bernoulli component  $g_j(X)$  in the best-fitting distribution as missing data, and uses expectation-maximisation (EM) to estimate the distribution of the missing data. Like SJPDA, there is a separate missing data distribution for each component  $a$  in

the multi-Bernoulli mixture; the distribution under the  $a$ -th component is  $q_a(\pi)$ . We constrain  $q_a(\pi) \geq 0; \forall a, \pi$ , and  $\sum_{\pi \in \Pi_N} q_a(\pi) = 1 \forall a$ . Accordingly, solution of (19) is equivalent to minimisation of  $J([g_j])$ , where

$$J([g_j]) = - \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) \cdot \log \left( \sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i) \right) \delta X_1 \cdots \delta X_N \quad (21)$$

$$= \sum_{a \in \mathcal{A}} w_a \left( \sum_{\pi \in \Pi_N} q_a(\pi) \right) \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) \cdot \log \left( \frac{\sum_{\pi \in \Pi_N} q_a(\pi)}{\sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i)} \right) \delta X_1 \cdots \delta X_N \quad (22)$$

$$\leq \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \int \cdots \int \prod_{i=1}^N f_{h_i}(X_i) \cdot \log \left( \frac{q_a(\pi)}{\prod_{i=1}^N g_{\pi(i)}(X_i)} \right) \delta X_1 \cdots \delta X_N \quad (23)$$

$$= \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \log q_a(\pi) - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{i=1}^N \int f_{h_i}(X_i) \log g_{\pi(i)}(X_i) \delta X_i \quad (24)$$

$$\leq T \cdot \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \log q_a(\pi) - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{i=1}^N \int f_{h_i}(X_i) \log g_{\pi(i)}(X_i) \delta X_i \quad (25)$$

$$\triangleq \tilde{J}_T([g_j], [q_a(\pi)])$$

where (22) simply multiplies by  $\sum_{\pi \in \Pi_N} q_a(\pi) = 1$  and similarly adds  $\log(1) = 0$ , (23) invokes the log-sum inequality, and (24) replaces the log of a product with the sum of logs and simplifies. This process is commonly used to understand the behaviour of EM, e.g., [14, p363]. In (25), we observe that the first term is negative (since  $0 \leq q_a(\pi) \leq 1$ ), hence incorporating a multiplier  $0 \leq T \leq 1$  loosens the bound. In statistical physics this corresponds to the temperature (e.g., [15]); the need for the change is discussed in section III-B.

**Problem 2.** BFMB may be solved approximately by minimising the upper bound to the objective of problem 1:

$$\operatorname{minimise}_{[g_j], [q_a(\pi)]} \tilde{J}_T([g_j], [q_a(\pi)]) \quad (26)$$

where  $q_a(\pi) \geq 0$ ,  $\sum_{\pi} q_a(\pi) = 1$ ,  $g_j(X) \geq 0$ ,  $\int g_j(X) \delta X = 1$ , and  $T \in [0, 1]$  is the temperature (a constant to be selected).

The standard method for solving the form of problem 2 is by block coordinate descent, alternating between minimisation with respect to  $[g_j]$  (M-step), and  $[q_a(\pi)]$  (E-step). These two

steps can be solved as:

$$g_j(X) = \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) f_{h_{\pi^{-1}(j)}}(X) \quad (27)$$

$$q_a(\pi) \propto \prod_{i=1}^N \exp \left\{ \frac{1}{T} \int f_{h_i}(X) \log g_{\pi(i)}(X) \delta X \right\} \quad (28)$$

If the distributions  $[g_j]$  are to be constrained to be Bernoulli-Gaussian, (27) is replaced by expressions matching the probability of existence, mean and covariance to the expression in (27). The Bernoulli-Gaussian form is convenient since it permits closed-form evaluation of (28). Specifically, if

$$f_h(X) = \begin{cases} 1 - r_h, & X = \emptyset \\ r_h \mathcal{N}\{x; \mu_h, \Sigma_h\}, & X = \{x\} \end{cases} \quad (29)$$

$$g_j(X) = \begin{cases} 1 - \hat{r}_j, & X = \emptyset \\ \hat{r}_j \mathcal{N}\{x; \hat{\mu}_j, \hat{\Sigma}_j\}, & X = \{x\} \end{cases} \quad (30)$$

then the M-step reduces to setting:

$$\hat{r}_j = \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) r_{h_{\pi^{-1}(j)}} \quad (31)$$

$$\hat{\mu}_j = \frac{1}{\hat{r}_j} \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) r_{h_{\pi^{-1}(j)}} \mu_{h_{\pi^{-1}(j)}} \quad (32)$$

$$\begin{aligned} \hat{\Sigma}_j &= \frac{1}{\hat{r}_j} \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) r_{h_{\pi^{-1}(j)}} \left\{ \Sigma_{h_{\pi^{-1}(j)}} \right. \\ &\quad \left. + [\mu_{h_{\pi^{-1}(j)}} - \hat{\mu}_j][\mu_{h_{\pi^{-1}(j)}} - \hat{\mu}_j]^T \right\} \end{aligned} \quad (33)$$

while the integral required in the E-step becomes:

$$\begin{aligned} - \int f_{h_i}(X) \log g_j(X) \delta X &= -(1 - r_{h_i}) \log(1 - \hat{r}_j) \\ &\quad - r_{h_i} \log \hat{r}_j + \frac{r_{h_i}}{2} \left\{ \text{trace}(\hat{\Sigma}_j^{-1} \Sigma_{h_i}) \right. \\ &\quad \left. + [\mu_{h_i} - \hat{\mu}_j]^T \hat{\Sigma}_j^{-1} [\mu_{h_i} - \hat{\mu}_j] + \log |2\pi \hat{\Sigma}_j| \right\} \end{aligned} \quad (34)$$

In most applications of EM, the missing data is estimated for each of a finite number of training samples. In turn, this guarantees that the log-sum inequality is tight at the optimum, and hence that the procedure will converge to a local minimum of the original likelihood function  $J([g_j])$ . In the present application, each point in the multi-target distribution  $f(X)$  is effectively a training sample, so to maintain a similar guarantee, the missing data would need to be permitted to be a function of the joint target state  $X = \bigcup_{i=1}^N X_i$ ; this was the major source of intractability in KLSJPDA [9]. The proposed method may be viewed as constraining the missing data to vary only with the global association hypothesis  $a$ , or simply minimising an upper bound that is not tight at the optimum. SJPDA simplifies the complexity of KLSJPDA similarly, constraining the assignment to depend only on the association hypothesis.

### B. Numerical demonstration

We demonstrate the approach on a simple one-dimensional case involving two Bernoulli-Gaussian tracks. The prior distributions have a probability of existence of 0.8 and 0.9, means

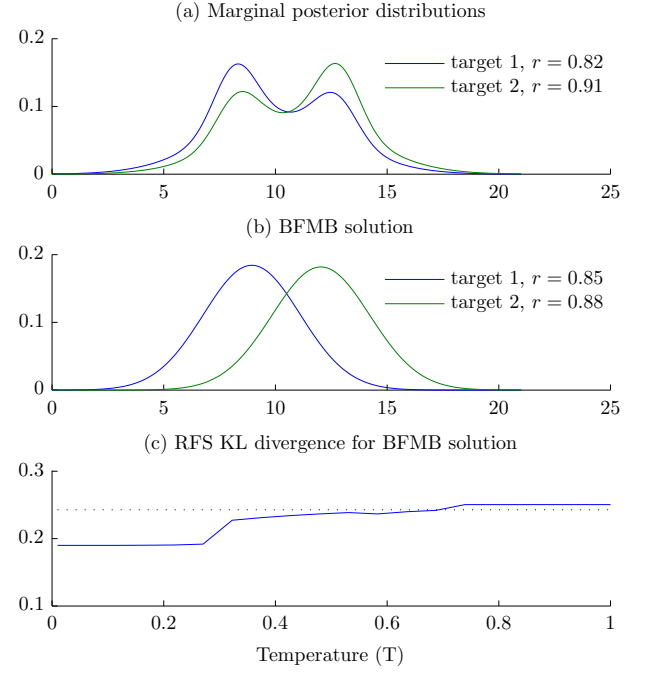


Fig. 1. Results of experiments: (a) shows the marginal posterior distribution, (b) shows the result of BFMB with  $T = 0.01$ , and (c) shows the RFS KL divergence (3) of the solution of BFMB as a function of the temperature  $T$  (evaluated numerically).

of 10 and 11, and a variance of 9. Two measurements are received, at coordinates 8 and 13. The probability of detection is 0.6, the false alarm intensity is 0.1, and the variance of the additive Gaussian measurement noise is 1.

The marginal posterior distribution of each track is shown in figure 1(a) (i.e., that obtained using TOM-MeMBer/P described in [2], conceptually similar to JPDA); the initialisation used for BFMB is a series of Bernoulli-Gaussian distributions matching the probability of existence, mean and covariance of these posterior marginals. The solution obtained using BFMB with  $T \approx 0$  is shown in figure 1(b). While a Gaussian distribution fit to the marginals would attempt to cover both modes (leading to two very similar tracks with much higher variance), BFMB successfully separates the two peaks, eliminating coalescence. The RFS KL divergence (3) of the BFMB solution is shown in figure 1(c) (evaluated numerically). The dotted line is the KL divergence of the Bernoulli-Gaussian distribution fit to the marginal posterior (i.e., the initialisation used in the optimisation). With  $T > 0.7$ , the algorithm converges (very slowly) to two identical Gaussian distributions covering both modes. This occurs because the reduction in the bound  $\tilde{J}_T$  obtained by setting  $q_a(\pi)$  to uniform (thereby maximising its entropy) outweighs the improvement that can be achieved in the mean log-likelihood of the true data under the simplified distribution. When the upper bound is loosened by reducing the temperature (reducing the influence of the entropy term), the algorithm focuses on improving the log-likelihood of the simplified distribution, and convergence occurs rapidly. The result in figure 1(c) demonstrates that the solutions obtained in this case have a lower RFS KL divergence (i.e., the original objective which could not be tractably optimised).

### C. Zero temperature case

Section III-B showed that the best performance in terms of the original objective (problem 1) occurs when we set  $T = 0$ . In this case, E-step reverts to a LP, which can be implemented through finding the most likely assignment  $\pi_a$  for each association hypothesis  $a \in \mathcal{A}$  using methods such as the auction algorithm. However, a deeper consideration of the geometry of the LP reveals a family of potential approximations.

**Theorem 2.** *Problem 2 can be solved equivalently as:*

$$\begin{aligned} \underset{q(h,j) \in \mathcal{P}}{\text{minimise}} & - \sum_{j=1}^N \int \left( \sum_{h \in \mathcal{H}} q(h,j) f_h(X) \right) \\ & \cdot \log \left( \sum_{h \in \mathcal{H}} q(h,j) f_h(X) \right) \delta X \end{aligned} \quad (35)$$

where the polytope  $\mathcal{P}$  is:

$$\mathcal{P} = \left\{ q(h,j) = \sum_{i=1}^N \left( \sum_{a=(h_1, \dots, h_N) \in \mathcal{A} | h_i=h} w_a \sum_{\pi \in \Pi_N | \pi(i)=j} q_a(\pi) \right) \right. \\ \left. \left| q_a(\pi) \geq 0, \sum_{\pi \in \Pi_N} q_a(\pi) = 1 \right. \right\} \quad (36)$$

*Proof:* To commence, consider the result obtained by substituting (27) into (25). In this case, when  $T = 0$ , the problem becomes:

$$\begin{aligned} \tilde{J}^*([q_a(\pi)]) &= \min_{[g_j]} \tilde{J}_T([g_j], [q_a(\pi)]) \\ &= - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{i=1}^N \int f_{h_i}(X) \cdot \\ & \cdot \log \left( \sum_{\tilde{a} \in \mathcal{A}, \tilde{\pi} \in \Pi_N} w_{\tilde{a}} q_{\tilde{a}}(\tilde{\pi}) f_{h_{\tilde{\pi}^{-1}(\pi(i))}}(X) \right) \delta X \end{aligned} \quad (37)$$

$$\begin{aligned} &= - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{j=1}^N \int f_{h_{\pi^{-1}(j)}}(X) \cdot \\ & \cdot \log \left( \sum_{\tilde{a} \in \mathcal{A}, \tilde{\pi} \in \Pi_N} w_{\tilde{a}} q_{\tilde{a}}(\tilde{\pi}) f_{h_{\tilde{\pi}^{-1}(j)}}(X) \right) \delta X \end{aligned} \quad (38)$$

$$\begin{aligned} &= - \sum_{j=1}^N \int \left( \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) f_{h_{\pi^{-1}(j)}}(X) \right) \cdot \\ & \cdot \log \left( \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) f_{h_{\pi^{-1}(j)}}(X) \right) \delta X \end{aligned} \quad (39)$$

$$\begin{aligned} &= - \sum_{j=1}^N \int \left( \sum_{h \in \mathcal{H}} q(h,j) f_h(X) \right) \cdot \\ & \cdot \log \left( \sum_{h \in \mathcal{H}} q(h,j) f_h(X) \right) \delta X \end{aligned} \quad (40)$$

where

$$q(h,j) \triangleq \sum_{i=1}^N \left( \sum_{a=(h_1, \dots, h_N) \in \mathcal{A} | h_i=h} w_a \sum_{\pi \in \Pi_N | \pi(i)=j} q_a(\pi) \right)$$

The step (38) changes the variable of summation from  $i$  to  $j = \pi(i)$  (noting that  $\pi$  is a bijection). ■

The objective (35) is the sum of entropies of the simplified Bernoulli components, which was proposed as a heuristic in [16]. Substituting the expression for  $g_j(X)$  back into (40), we return to a form that can be minimised via coordinate descent:

$$\begin{aligned} &\tilde{J}_T([g_j], q(h,j)) \\ &= - \sum_{j=1}^N \int \left( \sum_{h \in \mathcal{H}} q(h,j) f_h(X) \right) \log g_j(X) \delta X \end{aligned} \quad (41)$$

### D. Efficient approximation

Whereas the optimisation in the statement of problem 2 involves missing data  $q_a(\pi)$  for every global association hypothesis  $a \in \mathcal{A}$ , theorem 2 provides a form of the objective which depends only on the vastly simplified representation  $q(h,j)$ , specifying the weight of single target hypothesis  $h \in \mathcal{H}$  in the new Bernoulli component  $g_j(X)$ . This does not necessarily reduce complexity, however, as the feasible set  $\mathcal{P}$  suffers from combinatorial complexity. It does, however, raise the prospect of tractable approximations based on relaxations of the polytope  $\mathcal{P}$  that admit a compact description.

It is clear that any set of distributions  $[q_a(\pi)]$  will yield  $q(h,j)$  which satisfies the following constraints:

$$\sum_{h \in \mathcal{H}} q(h,j) = 1 \quad \forall j \in \{1, \dots, N\} \quad (42)$$

$$\sum_{j=1}^N q(h,j) = p_h \quad \forall h \in \mathcal{H} \quad (43)$$

where we assume that  $\sum_{a \in \mathcal{A}} w_a = 1$ , and

$$p_h = \sum_{i=1}^N p_i(h) \quad (44)$$

$$p_i(h) = \sum_{a=(h_1, \dots, h_N) \in \mathcal{A} | h_i=h} w_a \quad (45)$$

This suggests a relaxation of the polytope  $\mathcal{P}$  to the compact approximation:

$$\mathcal{M} = \left\{ q(h,j) \geq 0 \left| \sum_{h \in \mathcal{H}} q(h,j) = 1 \quad \forall j \in \{1, \dots, N\}, \right. \right. \\ \left. \left. \sum_{j=1}^N q(h,j) = p_h \quad \forall h \in \mathcal{H} \right. \right\} \quad (46)$$

The notation  $\mathcal{M}$  is chosen due to its similarity to the marginal polytope approximation that is widely used in variational inference [10]. Many other approximations are possible, and approximations can be improved incrementally (e.g., [17]).

The details of the resulting algorithm are shown in figure 2; we refer to the method as variational MB (VMB). Note that

```

1: procedure VMB( $N, \mathcal{H}, [p_i(h)], [r_h], [\mu_h], [\Sigma_h]$ )
2:    $q(h, j) \leftarrow p_j(h) \forall h \in \mathcal{H}, j \in \{1, \dots, N\}$ 
3:    $p_h \leftarrow \sum_{i=1}^N p_i(h) \forall h \in \mathcal{H}$ 
4:   repeat
5:     Calculate  $[\hat{r}_j], [\hat{\mu}_j], [\hat{\Sigma}_j]$  using (31)-(33)
6:     Calculate  $C(h, j)$  using (34)  $\forall h, j$ 
7:     Solve the LP:
       minimise  $\sum_{h \in \mathcal{H}} \sum_{j=1}^N C(h, j) q(h, j)$  (47)
       subject to  $\sum_{h \in \mathcal{H}} q(h, j) = 1 \forall j$ 
                   $\sum_{j=1}^N q(h, j) = p_h \forall h$ 
                   $q(h, j) \geq 0 \forall h, j$ 
8:   until Sufficiently small progress is made in objective
       of LP from one iteration to next
9:   return  $[\hat{r}_j], [\hat{\mu}_j], [\hat{\Sigma}_j], q(h, j)$ 
10: end procedure

```

Fig. 2. Pseudo-code for VMB algorithm. The output of the algorithm can be taken as either the probabilities of existence, and covariances of the new Bernoulli-Gaussian components, or the weights  $q(h, j)$ , defining new Bernoulli-Gaussian mixture components  $g_j(X) = \sum_h q(h, j) f_h(X)$ .

we revert to a coordinate descent solution (minimising (41) over both  $[g_j]$  and  $q(h, j)$ ), and utilise the Bernoulli-Gaussian form detailed in (29)-(30) (finding the best-fitting Bernoulli-Gaussian form). Even if the desire is to return a Gaussian mixture representation, the Bernoulli-Gaussian form is recommended for two reasons. Firstly, it permits closed-form evaluation of the entropy, which is not otherwise possible.<sup>2</sup> Secondly (and more importantly), it incorporates the desire that Bernoulli components be made local in state space. In the case where the Bernoulli-Gaussian approximation is not made, there is no penalty for solutions which collect Gaussian mixture components that are arbitrarily dissimilar together in the same Bernoulli component. The algorithm is initialised with the marginal probabilities  $p_i(h)$ , as can be approximated efficiently using methods such as [18] (as described in [2]).

Finally, we note that the form of (47) is a network flow LP, referred to as a transportation problem [19]. Problems of this type admit rapid solution; we utilise a forward-reverse variant of the transport auction algorithm described in [20].

#### E. Numerical demonstration

We repeat the demonstration of section III-B for VMB. The resulting distributions are visually indistinguishable to those in figure 1(b), and hence are not shown. The probabilities of existence of the resulting Bernoulli components are  $\hat{r}_1 = 0.83$  and  $\hat{r}_2 = 0.9$ . The numerically evaluated KL divergence of the solution is 0.1894, as opposed to 0.19 for BFMB.

#### IV. APPROXIMATE MINIMUM MEAN OSPA ESTIMATION

The second problem we consider is that of finding the minimum mean OSPA (MMOSPA) estimate. We modify the

standard OSPA measure by omitting the leading  $\frac{1}{n}$  factor so that the measure *adds* over targets rather than *averaging* over targets.

**Problem 3.** Find  $\hat{X}$  that solves the optimisation

$$\operatorname{argmin}_{\hat{X}} \int d(X, \hat{X})^p f(X) \delta X \quad (48)$$

where if  $X = \{x_1, \dots, x_n\}$ ,  $Y = \{y_1, \dots, y_m\}$  and  $n \geq m$ ,

$$d(X, Y) \triangleq \left[ \min_{\pi \in \Pi_n} \sum_{j=1}^m d_c(x_{\pi(j)}, y_j)^p + c^p(n - m) \right]^{\frac{1}{p}} \quad (49)$$

We formulate the estimate as  $\hat{X} = \bigcup_{j=1}^N \hat{X}_j$ , where for each  $j$ ,  $\hat{X}_j$  is Bernoulli, i.e., either  $\hat{X}_j = \emptyset$  or  $\hat{X}_j = \{\hat{x}_j\}$ . Applying corollary 1, we arrive at an equivalent form of problem 3.

**Problem 4.** Find Bernoulli sets  $\hat{X}_j$ ,  $j \in \{1, \dots, N\}$ , that solve the optimisation

$$\operatorname{argmin}_{[\hat{X}_j]} \sum_{a \in \mathcal{A}} w_a \int \dots \int \prod_{i=1}^N f_{h_i}(X_i) d([X_i], [\hat{X}_j])^p \delta X_1 \dots \delta X_N \quad (50)$$

where

$$d([X_i], [\hat{X}_j]) = \min_{\pi \in \Pi_N} \left[ \sum_{i=1}^N d(X_i, \hat{X}_{\pi(i)})^p \right]^{\frac{1}{p}} \quad (51)$$

Note that for Bernoulli sets, (49) evaluates to:

$$d(X_i, \hat{X}_j) = \begin{cases} 0, & X_i = \hat{X}_j = \emptyset \\ c, & X_i = \emptyset, \hat{X}_j \neq \emptyset \text{ or } X_i \neq \emptyset, \hat{X}_j = \emptyset \\ d_c(x_i, \hat{x}_j), & X_i = \{x_i\}, \hat{X}_j = \{\hat{x}_j\} \end{cases} \quad (52)$$

A related problem is the following, which we will subsequently show to be closely related to OSPA.

**Problem 5.** Find Bernoulli sets  $[\hat{X}_j]$  that solve the optimisation

$$\operatorname{argmin}_{[\hat{X}_j]} \sum_{a \in \mathcal{A}} w_a \int \dots \int \prod_{i=1}^N f_{h_i}(X_i) s_\gamma([X_i], [\hat{X}_j])^p \delta X_1 \dots \delta X_N \quad (53)$$

where  $s_\gamma([X], [\hat{X}])$  is the softmax approximation of the OSPA distance:

$$s_\gamma([X_i], [\hat{X}_j]) = \left\{ \frac{-1}{\gamma} \log \sum_{\pi \in \Pi_N} \exp \left[ -\gamma \sum_{i=1}^N d(X_i, \hat{X}_{\pi(i)})^p \right] \right\}^{\frac{1}{p}} \quad (54)$$

Problem 5 replaces the minimum in the OSPA definition with the function log-sum-exp, which is commonly referred to as *softmax*; this function is illustrated in figure 3. Note that the definition of  $s_\gamma$  implicitly yields a function taking set arguments  $s_\gamma(X, \hat{X})$  for  $|X| \leq N$ ,  $|\hat{X}| \leq N$  via corollary 1.

<sup>2</sup>This could be overcome by introducing additional missing data describing the correspondence of mixture components, but this would further harm tractability.

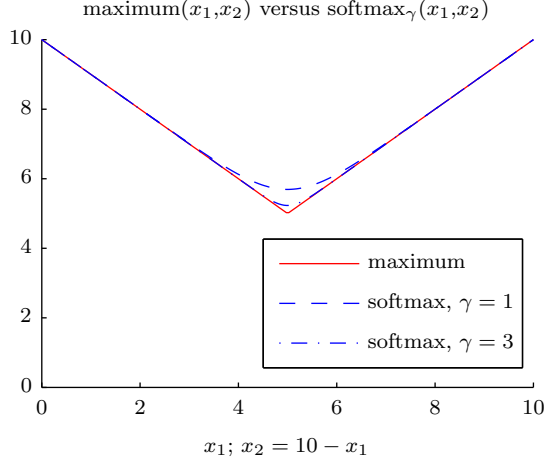


Fig. 3. Softmax function,  $\text{softmax}_\gamma(X) \triangleq \frac{1}{\gamma} \log \sum_{x \in X} \exp\{\gamma x\}$ .

#### A. Solution of MMOSPA via VMB

Note the similarity between (19) and problem 5. Specifically, with  $p = 2$ , if we set for  $j$  with  $\hat{X}_j = \{\hat{x}_j\}$

$$g_j(X) = \begin{cases} \exp\{-\gamma c^2\}, & X = \emptyset \\ \exp\{-\gamma d(x, \hat{x}_j)^2\} + \exp\{-\gamma c^2\}, & X = \{x\} \end{cases} \quad (55)$$

and for  $j$  with  $\hat{X}_j = \emptyset$ ,

$$g_j(X) = \begin{cases} 1, & X = \emptyset \\ \exp\{-\gamma c^2\}, & X = \{x\} \end{cases} \quad (56)$$

then the objective of problem 1 corresponds to that of problem 5.<sup>3</sup> Incorporating both (55) and (56) into a single expression via the parameterisation  $(\hat{r}_j, \hat{x}_j)$  where  $\hat{r}_j \in [0, 1]$ ,  $\hat{X}_j = \emptyset$  if  $\hat{r}_j = 0$ ,  $\hat{X}_j = \{\hat{x}_j\}$  if  $\hat{r}_j = 1$ , we finally obtain:

$$g_j(X; \hat{r}_j, \hat{x}_j) = \begin{cases} \exp\{-\gamma c^2\}^{\hat{r}_j}, & X = \emptyset \\ [\exp\{-\gamma d(x, \hat{x}_j)^2\} + \exp\{-\gamma c^2\}]^{\hat{r}_j} \cdot \exp\{-\gamma c^2\}^{1-\hat{r}_j}, & X = \{x\} \end{cases} \quad (57)$$

Subsequently, we relax the feasible set to  $\hat{r}_j \in [0, 1]$ , substitute (57) into (41), and repeat the EM derivation to incorporate additional missing data  $q_{h,j}(b)$ ,  $b \in \{0, 1\}$ , in order to handle the two terms in the second case of (55) (as in (22) and (23)). We also divide through by  $\gamma$ , and incorporate a smoothing term  $\frac{1}{\gamma} \sum_j [\hat{r}_j \log \hat{r}_j + (1 - \hat{r}_j) \log(1 - \hat{r}_j)]$  in order to gradually converge on integral values of  $\hat{r}_j$  as  $\gamma \rightarrow \infty$ .<sup>4</sup> The resulting

<sup>3</sup>It is out by a multiplicative constant  $\frac{1}{\gamma}$ , which we subsequently incorporate (although its absence would not affect the location of the minima). Note also that  $d_c(x_i, \hat{x}_j)$  is replaced by its softmax approximation in (55).

<sup>4</sup>Without the smoothing term, the first optimisation of  $\hat{r}_j$  would produce integral values, and if  $\hat{r}_j = 0$ , it would remain so for all subsequent iterations.

modified objective is:

$$\begin{aligned} \check{J}([\hat{r}_j], [\hat{x}_j], [q_{h,j}(b)], q(h, j)) = & \sum_{j=1}^N \sum_{h \in \mathcal{H}} q(h, j) \left\{ c^2(1 - r_h)\hat{r}_j + c^2 r_h(1 - \hat{r}_j) \right. \\ & + \frac{1}{\gamma} r_h \hat{r}_j \sum_{b=0}^1 q_{h,j}(b) \log q_{h,j}(b) \\ & + r_h \hat{r}_j \left[ q_{h,j}(0)c^2 + q_{h,j}(1) \int f_h(x) d(x, \hat{x}_j)^2 dx \right] \Big\} \\ & + \frac{1}{\gamma} \sum_{j=1}^N \{ \hat{r}_j \log \hat{r}_j + (1 - \hat{r}_j) \log(1 - \hat{r}_j) \} \end{aligned} \quad (58)$$

This can be optimised by block coordinate descent using the iterative equations:

$$q_{h,j}(b) \propto \begin{cases} \exp(-\gamma c^2), & b = 0 \\ \exp(-\gamma \int f_h(x) d(x, \hat{x}_j)^2 dx), & b = 1 \end{cases} \quad (59)$$

$$q(h, j) = \text{solution of the LP (47) using } C(h, j) \text{ in (60)}$$

$$\begin{aligned} C(h, j) = & \frac{1}{\gamma} r_h \hat{r}_j \sum_{b=0}^1 q_{h,j}(b) \log q_{h,j}(b) \\ & + c^2[(1 - r_h)\hat{r}_j + r_h(1 - \hat{r}_j) + r_h \hat{r}_j q_{h,j}(0)] \\ & + r_h \hat{r}_j q_{h,j}(1) \int f_h(x) d(x, \hat{x}_j)^2 dx \end{aligned} \quad (60)$$

$$\hat{x}_j = \frac{\sum_{h \in \mathcal{H}} q(h, j) r_h q_{h,j}(1) \mu_h}{\sum_{h \in \mathcal{H}} q(h, j) r_h q_{h,j}(1)} \quad (61)$$

$$\hat{r}_j = \frac{\alpha_j}{\alpha_j + \beta_j} \quad (62)$$

$$\begin{aligned} \alpha_j = & \exp \left\{ - \sum_{h \in \mathcal{H}} q(h, j) \left[ \gamma c^2(1 - r_h + r_h q_{h,j}(0)) \right. \right. \\ & + r_h \sum_{b=0}^1 q_{h,j}(b) \log q_{h,j}(b) \\ & \left. \left. + \gamma r_h q_{h,j}(1) \int f_h(x) d(x, \hat{x}_j)^2 dx \right] \right\} \end{aligned} \quad (63)$$

$$\beta_j = \exp \left\{ - \gamma c^2 \sum_{h \in \mathcal{H}} q(h, j) r_h \right\} \quad (64)$$

In the Bernoulli-Gaussian case,  $\int f_h(x) d(x, \hat{x}_j)^2 dx = \|\mu_h - \hat{\mu}_j\|^2 + \text{trace} \Sigma_h$ . As we gradually increase  $\gamma \rightarrow \infty$ , both  $\hat{r}_j$  and  $q_{h,j}(b)$  converge to integral solutions (as the softmax approximation converges to the maximum function).

The method is referred to as variational MMOSPA (VM-MOSPA).

#### B. Numerical demonstration

Again, the example from section III-B is used to demonstrate operation. The estimates  $\hat{x}_j$  and  $\hat{r}_j$  are initialised with the mean and probability of existence of the marginal distribution (shown in figure 1(a)). Initially,  $\gamma = 0.1$ , and it is multiplied by 10 after each iteration. With  $c = 4$ , the resulting estimate is  $\{8.8168, 12.0905\}$  (convergence is



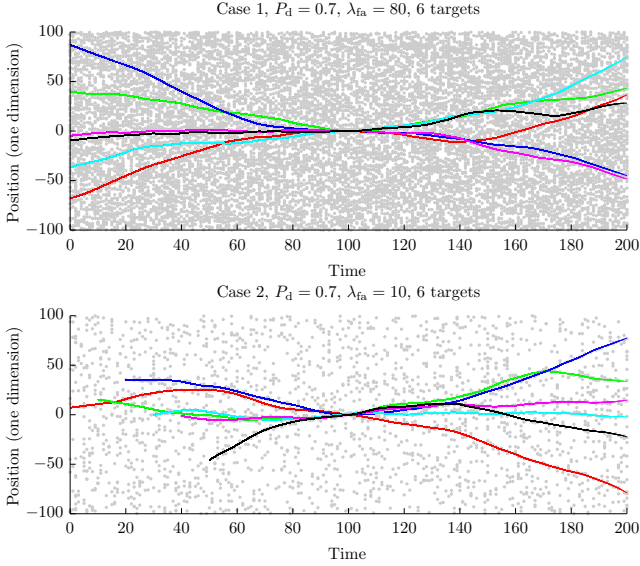


Fig. 4. One dimension of a single Monte Carlo run of the scenario cases 1 and 2, adopted from [2]. Target trajectories are shown in colours, and measurements are shown in grey. Both use  $P_d = 0.7$ ; top and bottom have the expected number of false alarms set to 80 and 10 respectively.

attained after three iterations). With  $c = 1$ , the resulting estimate is  $\{12.7621\}$ .

## V. EXPERIMENTS

In order to evaluate the performance of the proposed methods in a challenging scenario, we utilise the experiments from [2]. The scenarios involve  $n \in \{6, 10, 20\}$  targets which are in close proximity at the mid-point of the simulation, achieved by initialising at the mid-point and running forward and backward dynamics. We consider two cases for the mid-point initialisation (i.e.,  $t = 100$ ):

$$\text{Case 1: } x_{100} \sim \mathcal{N}\{0, 10^{-6} \times \mathbf{I}_{4 \times 4}\}$$

$$\text{Case 2: } x_{100} \sim \mathcal{N}\{0, 0.25 \times \mathbf{I}_{4 \times 4}\}$$

where the target state is position and velocity in two dimensions. Snapshots of one dimension of both cases are shown in figure 4. Case 1 represents a worst-case scenario for coalescence, since targets are completely indistinguishable (in position and velocity) at the mid-point. In case 2, there is a discernible difference in velocity, hence the effect is expected to be somewhat reduced. In case 1, targets all exist throughout the simulation (tracks are not pre-initialised). In case 2, the targets are born at times  $\{0, 10, \dots, 10(n-1)\}$  (any targets not existing prior to time  $t = 100$  are born at that time; consequently, for case 2 with  $n = 20$ , ten targets are born at time  $t = 100$ ).

Target-originated measurements provide position corrupted by Gaussian noise. Cases are considered with the expected number of false alarms per scan as  $\lambda_{fa} \in \{10, 20, 40, 60, 80\}$ , and with  $P_d \in \{0.3, 0.5, 0.7, 0.98\}$ , representing a range of SNR values. Further details of the simulations can be found in [2].

Further details of the implementation are provided in [2]. Marginal association probabilities (e.g., (45)) are calculated approximately using the variational method of [18]. The recycling method of [21] is applied to Bernoulli components with

a probability of existence less than 0.1. The VMB algorithm is applied to clusters of MB components (tracks) which share measurements (i.e., any hypothesis in the track). Tracks in which all hypotheses have only a single update are excluded from VMB, as are tracks which are in isolation (i.e., are not in a cluster with any other track). Consequently, even when using VMB with a Gaussian approximation, a Gaussian mixture representation is maintained for tracks in isolation. The MOM-MeMber/P algorithm extracts estimates by determining the mode of the cardinality distribution ( $\hat{n}$ ), and outputting the most likely hypothesis of the  $\hat{n}$  Bernoulli components with the highest probability of existence. Remaining methods output the most likely hypothesis of each Bernoulli component with a probability of existence of at least 0.8.

The scenarios examined are exceptionally challenging due to the large number of targets in close proximity. While others have considered larger numbers of targets, these are generally positioned uniformly in space and rarely come into close contact. Cases such as this can be effectively decoupled into series of single target tracking problems. In the present study, up to 20 targets have effectively the same position and velocity at the mid-point in time, and the dependency between targets is inescapable.

The results are shown in figure 5. The large error suffered by TOM-MeMber/P (solid blue) commencing shortly after time 100 corresponds to coalescence, similar to that experienced by JPDA/JIPDA. This occurs when targets have been closely spaced, and begin to separate. TOM-MeMber/P maintains a Gaussian mixture for each Bernoulli component (track), and outputs its estimate as the mean of the highest weighted Gaussian mixture component. After targets have been closely spaced, the Gaussian mixture for each target contains components representing all targets. Consequently, the highest weight component for all tracks can fall on the same target, leaving all other targets without estimates. TOM-MeMber/P using the VMOSPAs estimate (dashed blue) resolves this difficulty to an extent. The effectiveness of the method is hampered by the poor initialisation used (i.e., the mean of each track); it could be improved by initialising using some form of clustering algorithm.

The results demonstrate the success of VMB in resolving the coalescence phenomenon. In almost all cases, VMB outperforms TOM-MeMber/P and MoM-MeMber/P (in [2], it was shown that MoM-MeMber/P globally outperforms CB-MeMber and CPHD). The goal of this work was to combine the superior performance of TOM-MeMber/P in problems involving well-spaced targets with the robustness of MoM-MeMber/P in problems involving closely spaced targets: this has been achieved.

The small difference between VMB with (magenta) and without (red) the Gaussian approximation is somewhat surprising. It is only in the case with  $P_d = 0.3$  that there is a discernible difference. As previously noted, VMB (and hence the Gaussian approximation) is only applied when targets are in close proximity, so in a significant part of the simulation the methods both maintain Gaussian mixtures.

When applied alongside algorithms other than TOM-MeMber/P, VMOSPAs results in a small improvement in

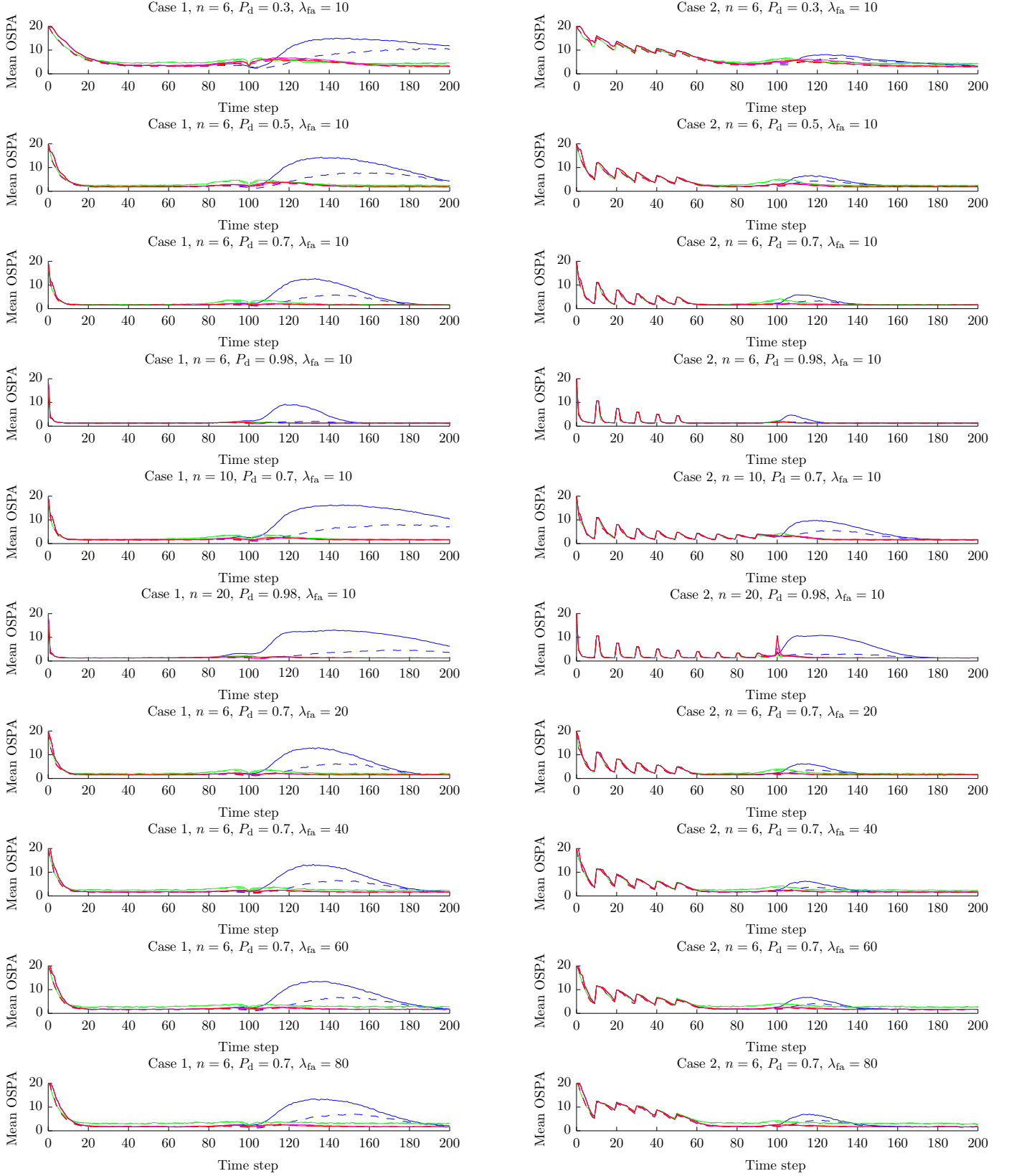


Fig. 5. Results of scenario. TOM-MeMber/P (similar to JIPDA) shown in blue, MOM-MeMber/P (similar to MeMber/CB-MeMber) in green, VMB in red (retaining a Gaussian mixture representation of Bernoulli components), and the Gaussian approximation of VMB in magenta. Each algorithm combined with the VMOSPA estimator shown with a dashed line of the corresponding colour.

TABLE I  
AVERAGE EXECUTION TIME FOR A COMPLETE MC TRIAL (IN SECONDS)  
FOR TOM-MEMBER/P, MOM-MEMBER/P, VMB WITH GAUSSIAN  
APPROXIMATION AND VMB RETAINING GAUSSIAN MIXTURE  
REPRESENTATION.

Case, $n$	$P_d, \lambda_{fa}$	TOMB	MOMB	VMBG	VMB
1, 6	0.98, 10	9	9	3	18
1, 6	0.7, 10	23	24	4	74
1, 6	0.5, 10	29	27	5	84
1, 6	0.3, 10	35	32	10	68
1, 10	0.7, 10	54	51	7	217
1, 20	0.98, 10	122	117	48	529
1, 6	0.7, 80	42	43	14	92
2, 10	0.7, 10	26	25	8	112
2, 20	0.98, 10	41	37	48	177
2, 6	0.98, 10	5	4	2	7
2, 6	0.5, 10	14	14	5	35

performance. The difference is more noticeable in the early stages of the simulation, where VMMOSPA determines that it is beneficial to output estimates before the probability of existence exceeds the 0.8 threshold used by TOM-MeMBer/P and VMB.

The computation time for the methods is shown in table I. The table shows the average time to execute a complete Monte Carlo (MC) simulation (consisting of 201 time steps). The table demonstrates that, particularly with a Gaussian approximation, VMB is highly tractable. Without the Gaussian approximation, the time is considerably longer. The current implementation retains up to 2000 hypotheses (in total, across all targets), thus with 20 targets, it solves a  $20 \times 2000$  network flow optimisation at the mid-point of the simulation. The time could be reduced dramatically by applying the optimisation to a simplified representation of the Gaussian mixture, even if a more detailed representation will ultimately be retained. Details of this will be examined in future publications. Computation times for VMMOSPA (not shown) range from 34 sec (six targets,  $P_d = 0.98$ ) to 3041 sec (20 targets). The current implementation of VMMOSPA does not perform clustering, hence it faces a very large optimisation at every time step. This could also be reduced to something less than VMB through appropriate application of clustering and mixture reduction. Again, details of these computational optimisations will be the subject of future examinations.

## VI. CONCLUSION

This paper has presented a principled, highly efficient, approximate method for finding the MB distribution that minimises the KL divergence from the full RFS distribution. To date, there have been two practical difficulties that have limited application of the JPDA/JIPDA family of trackers to problems involving closely-spaced targets. The first is the intractability of the calculation of marginal association probabilities (e.g., (45)). A highly accurate approximation of these quantities based on variational methods was examined in [18]. The second limitation was the problem of coalescence, which has been tractably addressed in this paper. Consequently, this work represents a significant step forward in the practical applicability of JPDA and related methods.

## REFERENCES

- [1] J. L. Williams, "Experiments with graphical model implementations of multiple target multiple Bernoulli filters," in *Proc. 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Adelaide, Australia, December 2011, pp. 532–537.
- [2] —, "Marginal multi-Bernoulli filters (extended version)," *Submitted to IEEE Trans. Aerosp. Electron. Syst.*, 2013. [Online]. Available: <http://arxiv.org/abs/1203.2995v2>
- [3] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3460–3475, 2013.
- [4] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE J. Ocean. Eng.*, vol. 8, no. 3, pp. 173–184, Jul 1983.
- [5] D. Musicki and R. J. Evans, "Joint integrated probabilistic data association: JPDA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1093–1099, July 2004.
- [6] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.
- [7] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 409–423, Feb. 2009.
- [8] H. A. Blom and E. A. Bloem, "Probabilistic data association avoiding track coalescence," *IEEE Trans. Autom. Control*, vol. 45, no. 2, pp. 247–259, February 2000.
- [9] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multitarget tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4677–4691, October 2011.
- [10] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [11] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, August 2008.
- [12] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting two birds with two bullets: how to find minimum mean OSPA estimates," in *Proc. 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.
- [13] D. Musicki, R. Evans, and S. Stankovic, "Integrated probabilistic data association," *IEEE Trans. Autom. Control*, vol. 39, no. 6, pp. 1237–1241, Jun 1994.
- [14] K. P. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT Press, 2012.
- [15] J. S. Yedidia, "An idiosyncratic journey beyond mean field theory," in *Advanced Mean Field Methods—Theory and Practice*, M. Opper and D. Saad, Eds. Cambridge, MA: MIT Press, 2001, pp. 21–35.
- [16] J. L. Williams, "Alternative multi-Bernoulli filters (extended version)," arXiv, e-print arXiv:1203.2995v1, March 2012. [Online]. Available: <http://arxiv.org/abs/1203.2995>
- [17] D. Sontag, "Approximate inference in graphical models using lp relaxations," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2010.
- [18] J. L. Williams and R. A. Lau, "Approximate evaluation of marginal association probabilities with belief propagation," *Submitted to IEEE Trans. Aerosp. Electron. Syst.*, 2012. [Online]. Available: <http://arxiv.org/abs/1209.6299>
- [19] D. Bertsekas, *Network Optimization: Continuous and Discrete Models*. Belmont, MA: Athena Scientific, 1998. [Online]. Available: <http://mit.edu/dimitrib/www/net.html>
- [20] D. P. Bertsekas and D. A. Castañón, "The auction algorithm for the transportation problem," *Annals of Operations Research*, vol. 20, no. 1, pp. 67–96, December 1989.
- [21] J. L. Williams, "Hybrid Poisson and multi-Bernoulli filters," in *Proc. 15th International Conference on Information Fusion*, Singapore, July 2012, pp. 1103–1110.